Multimodal Command Interaction:
Scientific and Technical Final Report

November 25, 2003

Sponsored by

Defense Advanced Research Projects Agency (DOD)
(Information Exploitation Office)

ARPA Order K475/40

Issued by U.S. Army Aviation and Missile Command Under

Contract No. DAAH01-02-C-R051

Name of Contractor: Natural Interaction Systems, LLC.
Principal Investigator: Philip R. Cohen, Ph.D.
Business Address: 10260 SW Greenburg Rd.
Portland, OR 97223
Phone Number: 503-293-8414

Effective Date of Contract: October 31, 2001
Reporting Period: October 31, 2001 through November 25, 2003

DISCLAIMER

"The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Defense Advanced Research Projects Agency or the U.S. Government"

Approved for public release; distribution unlimited.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (Leave Blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	November 25, 2003	Scientific and Technical Final Report, October 31, 2001 through November 25, 2003

4. TITLE AND SUBTITLE
Multimodal Command Interaction

5. FUNDING NUMBERS
C: DAAH01-02-C-R051

6. AUTHOR(S)
Philip R. Cohen, Ph.D.
David R. McGee, Ph.D.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Natural Interaction Systems, LLC.
10260 SW Greenburg Rd., Suite 400
Portland, OR 97223

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Defense Advanced Research Projects Agency (DOD)
Mr. Ward Page, Information Exploitation Office
701 North Fairfax Drive
Arlington, VA 22203-1714

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES
None.

12a. DISTRIBUTION/AVAILABILITY STATEMENT (see Section 5.3b of this solicitation)

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

This is the final report for DARPA SBIR No. DAAH01-02-C-R051. The objective of this Phase I SBIR was to demonstrate, through the development of a proof-of-concept, that a combined paper/digital system for capturing, sharing, and understanding command, control, communications, computers, intelligence, surveillance, and reconnaissance (C4ISR) information sketched on paper maps was feasible. In this final report, we discuss

1. an introduction to the set of C4ISR problems in command-posts and elsewhere that such technology would aid in solving,
2. background research that our small business, Natural Interaction Systems, LLC., brings to bear in designing and developing software solutions that enable multimodal interaction on various types of computing displays and architectures,
3. the designs and successful implementations of these prototypes,
4. the demonstrations that have resulted in follow-on work.

The interim results of Phase I were so promising that an extension was granted, allowing us to add spoken language as another possible input modality when working with the paper maps. Ultimately, a Phase II contract was granted. This work will be featured in the forthcoming January special issue on Multimodal Interaction in the Communications of the ACM.

14. SUBJECT TERMS
C2, C4ISR, multimodal, human-computer interface, interaction, HCI, sketch, pen, voice

15. NUMBER OF PAGES
18

16. PRICE CODE
N/A

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL

Abstract

This is the final report for DARPA SBIR No. DAAH01-02-C-R051. The objective of this Phase I SBIR was to demonstrate, through the development of a proof-of-concept, that a combined paper/digital system for capturing, sharing, and understanding command, control, communications, computers, intelligence, surveillance, and reconnaissance (C4ISR) information sketched on paper maps was feasible. In this final report, we discuss

5. an introduction to the set of C4ISR problems in command-posts and elsewhere that such technology would aid in solving,

6. background research that our small business, Natural Interaction Systems, LLC., brings to bear in designing and developing software solutions that enable multimodal interaction on various types of computing displays and architectures,

7. the designs and successful implementations of these prototypes,

8. the demonstrations that have resulted in follow-on work.

The interim results of Phase I were so promising that an extension was granted, allowing us to add spoken language as another possible input modality when working with the paper maps. Ultimately, a Phase II contract was granted. This work will be featured in the forthcoming January special issue on Multimodal Interaction in the Communications of the ACM.

Table of Contents

Introduction: The unmet interface needs of warfighters

Although the graphical user interface (GUI) pervades military systems, it has long been recognized that this interface technology frequently does not meet the needs of warfighters. For example, GUIs do not support mobile users, users of small devices, or users whose hands and eyes are busy. Even in fixed facilities, users spend inordinate amounts of time manipulating menus and filling in forms, rather than thinking about command and control. Furthermore, the software tools provided often do not match the task – e.g., instead of allowing users to draw their symbology in a continuous fashion, users are required to click on various points on the map, after which the system "connects the dots." This technique makes nuanced sketching impossible, and prevents the user from being as precise as s/he may want to be. Although it may appear that despite these difficulties, the GUI still provides an effective baseline capability, we believe these problems are true impediments to the adoption of digital systems by warfighters.

At USMC bases, at Ft. Leavenworth, Kansas, during Army National Guard exercises, and elsewhere we observed commanders and their subordinates engaging in command and control of armed forces. The photograph in Figure 1 was taken during an especially frenetic period in an Army division command post during an exercise. At the left is a rear-projected SMART Board™ and at the right is a Sun Microsystems workstation. Several other systems, not captured in the photograph, are in the immediate foreground. On each display is one of the latest command and control software systems. Notice, however, that no one is using these systems during this critical phase of the operation.

Rather, the commander and his staff have chosen to use a set of tools other than the computerized ones designed for this task. They have quite purposefully turned their backs on computer-based tools and graphical user interfaces, preferring instead to employ an 8-foot-high by 6-foot-wide paper map, arrayed with Post-it notes (Figure 1). This use of paper does not just happen during exercises. Below (Figure 2) are photos from the Coalition Forces Land Component Command center (CFLCC) in Doha, Qatar during Operation Iraqi Freedom. These photos are taken in the "war room," where the decision-makers work, as opposed to the auditorium-sized operations center. Notable in these photographs are general officers using paper maps, some with Post-It notes.

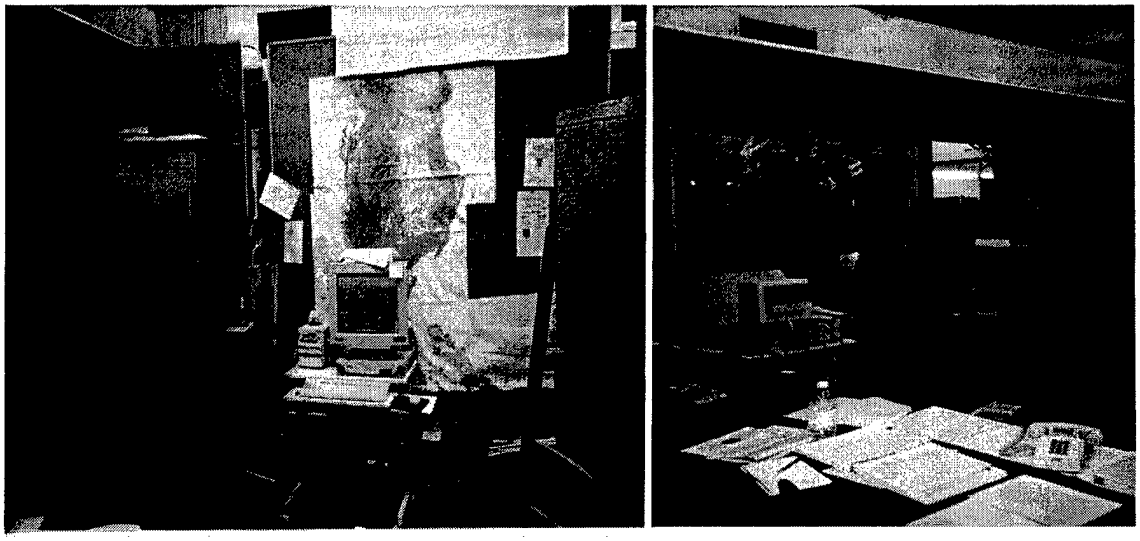

Figure 1: Left – A division command post during an exercise; Right – What operators actually use

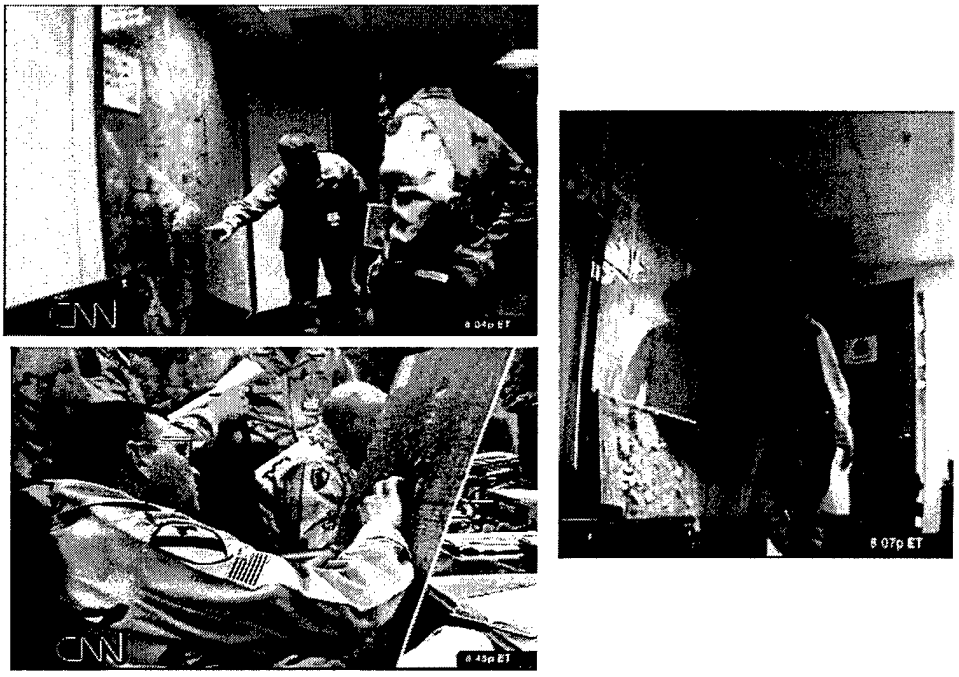

Figure 2. CFLCC War Room during OIF. Notice the use of paper maps

In general, it is fair to say that despite the best efforts of technologists and military research and development laboratories, many warfighters avoid computers.[1] However,

[1] They are not alone in this preference. It has been observed that doctors in emergency and intensive care facilities, as well as air traffic controllers, prefer to interact with physical objects, such as paper and pencil, rather than use a

the officers are not simply trying to be difficult or overly conservative in rejecting digital systems. Rather, the systems they have received are missing qualities that they value highly. The users tell us that they continue to use paper maps because they have extremely high resolution, are malleable, cheap, lightweight, and can be rolled up and taken anywhere. Importantly, paper does not fail, and it supports face-to-face collaboration among the staff members.

Thus, because of a variety of factors, officers may prefer paper to digital systems. We believe there is no reason they cannot have the benefits of both paper and digital technology. To prove this, we undertook Phase I SBIR research and development effort that developed the concept of collaborative, paper-based multimodal interfaces for military environments. In order to understand the relationship of the paper-based system to multimodal interaction, we provide background information on multimodal map-based interaction using speech and sketch.

Background

In this project, we developed and transition multimodal technology that enables commanders and their staff to interact with decision support systems through spoken language coupled with pen-based sketching and writing. This capability enables officers to create courses of action, initialize simulators, lay down forces on a map, and so forth. Multimodal interaction is advantageous for map-based tasks (Oviatt, 1996), for mobile users (Oviatt, 2000a and 2000b) and for diverse user populations for whom a one-size-fits-all user interface style complicates training and operations. Moreover, multimodal interaction supports more robust performance than unimodal systems since input from one mode can compensate for errors in another (McGee, et al. 1998; Oviatt, 1999). Based on prior DARPA, ONR, and NSF support, we have developed and empirically analyzed, both in the laboratory and the field, leading technology that supports such multimodal interaction using either touch-enabled LCD-based or paper-based (i.e., paper maps and Post-it notes) interfaces. These systems are called QuickSet (Cohen, et al. 1997) and Rasa (McGee, et al. 2000, 2002), respectively, and have been licensed exclusively to Natural Interaction Systems, LLC for commercialization.

QuickSet

QuickSet is a collaborative handheld multimodal system based on a multiagent architecture, which controls numerous applications, including simulators (ModSAF), exercise initialization (ExInit), and virtual terrain environments (the Naval Research Laboratory's Dragon II and Panda and SPAWAR's CommandVu). Most recently, it has been incorporated into NRL's BARS mobile augmented reality system, and now interoperates with Northrop Grumman's IBCP common operational picture for the ABCS

---

computer (Gorman, et al. 2000; Mackay, 1999). What is common among all these environments is their life-and-death nature, and the users' absolute requirement for safety and robustness. One key reason that air traffic was able to resume as quickly as it did after the evacuation of the SeaTac air traffic control center during the recent Seattle earthquake is that controllers are used to employing "flight strips" – handwritten records of incoming flights that are arrayed on a table to indicate relative altitudes. If controllers had to improvise a safe and robust manual method when they moved to temporary quarters equipped only with radios, the airport would undoubtedly have been shut down for a much longer time.

suite of C2 systems. QuickSet enables users to create units and control measures on a PC screen, using a variety of form factors ranging from handheld or wearable devices to wall-sized displays, simply by speaking and sketching. For example, the user can create and position an M1A1 company at a given location and with a given orientation and posture by saying: "M1A1 company facing one two zero degrees in defensive posture," while touching the desired location. In contrast, a user of a graphical user interface (GUI) would have to locate the desired unit in a browser or palette, drag the icon onto the screen, and fill in various parameters in a dialogue box. Likewise, by speaking and sketching, the user can create and label control measures, such as phase lines, and unit boundaries, as well as objectives, routes, fortifications, axes of advance, air corridors, no go/slow go areas, restricted fire areas, drop zones, supply routes, cultural features, etc. Creating control measures via a GUI is much more cumbersome than speaking and drawing, and usually involves selecting the control measure type from a menu, and touching/clicking on the map to enter individual points that will be connected. Such interfaces are particularly awkward and error prone when users are mobile (c.f., the Urban Warrior exercise), and when the device is either very small or very large. Figure 3 show photos of QuickSet in use with a handheld tablet PC (left) and a 50" plasma display with touch panel (right). Note that the systems are collaborating.

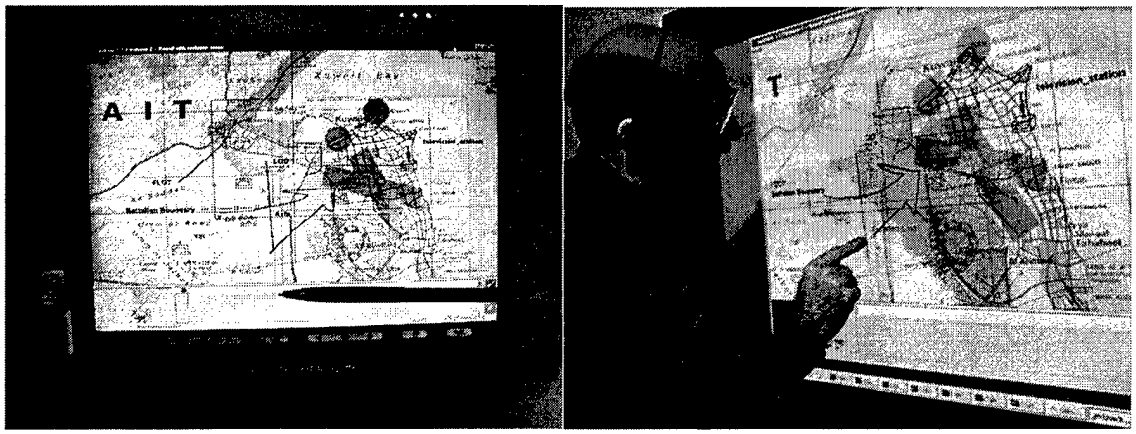

Figure 3: Left: QuickSet operating on a handheld PC; Right: Collaborating QuickSet operating on a 50" plasma display with touch overlay. User is speaking through a wireless microphone while drawing an axis of advance with his finger.

The QuickSet user creates entities on various layers in the local system's database. When another user connects to the database on the local machine, the entities are then shared. Likewise, QuickSet's multiagent system facilitator sends object update messages to subscribing agents. Such a capability enables multiple user interfaces to collaborate, by enabling them to couple so that panning and/or zooming of one does the same to the other, etc. The database can populate and control other digital systems, and can interoperate with DII/COE systems via CORBA, and ABCS systems via XML. Other interoperation capabilities include a bridge to the CoABS Grid, which enabled QuickSet to interoperate with systems from twenty contractors during a recent demonstration. The ultimate goal for QuickSet development is to build a multimodal "Battleboard" – an officer's sketchpad that supports rapid course of action (COA) creation, simulation, visualization, and collaboration.

Within QuickSet (see Figure 4), multiple streams of information flow via the Adaptive Agent Architecture (Kumar, 2000) to independent recognizers and are fused at the level of meaning. Information can come at any time, with or without accompanying information in another modality. Based on this architecture, we were able to build a first prototype paper-based system, in which we decoupled the display from the underlying digitizing of input. Instead of a computer display, we used something much higher in resolution, lighter in weight, and more robust — a paper map. This multimodal system, Rasa, was able to demonstrate the advantages of paper and of digital systems. Below, we describe how this works.

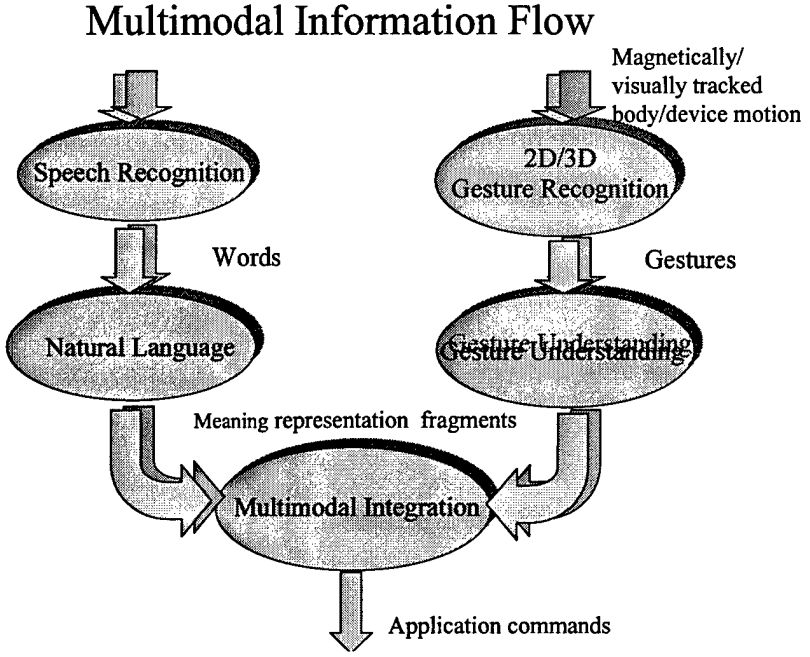

Figure 4: Generic multimodal information flow

Rasa

Rasa enables a military officer to use paper maps and Post-It™ notes (McGee, et al. 2000) in support of command and control tasks. During battle tracking, officers plot unit locations, activities, and other elements on a paper map by drawing unit symbols on Post-It notes, positioning them on the map, and then moving them in response to incoming reports of their new locations.[2] With Rasa (Figure 5), each of the pieces of paper is mounted on a digitizing tablet — the map is registered to a large touch-sensitive tablet, and the Post-Its rest upon a tablet that supports both digital and physical ink. The user writes a military unit symbol (e.g., for an armored platoon) on a Post-It note. The user can also speak identifying information about that unit (e.g., "First platoon, Charlie Company") while drawing the symbol. The computer recognizes both the symbol and

[2] Even with modern GPS-based Blue Force tracking capabilities, the tracking of red entities may still require some hand-entry, either at the point of observation, or in the CP.

the voice, fusing their meanings using multimodal integration techniques (Cohen, et al. 1997). Then, the user places the Post-It onto the paper map, which causes the unit that was drawn (and sometimes spoken about) to be recorded in the system's database, with its location specified by the place it was mounted on the map. The system projects a unit symbol onto the relevant location on the paper map, and distributes the result to collaborating systems. If a report arrives indicating that the unit has moved, the user only needs to pick up the note and put it at the new location on the map.

In military environments, work must somehow continue during a system or communications failure. What happens if the computer supporting Rasa goes down? In order to investigate this question, we undertook an experiment (McGee et al., 2002) in which officers were studied as they tracked a battle. During each session, Rasa was deliberately "crashed," but reports kept arriving. In response, officers simply continued to create and move the Post-It notes on the paper map. When the computer came back online, it digitally projected the old unit locations onto the paper map, whereas the physical notes indicated the units' current locations. It was then a simple process to reconcile the computer system with the paper version. Thus, because the physical objects constituted the user interface, no additional ongoing "backup" process was needed. Whereas Rasa enables users to employ paper maps in performing their tasks, the maps still require large, relatively immobile digitizers. Moreover, because of limitations in current digitizing tablets, only one person can write on the map at a time. Thus, Rasa is still limited in its mobility and support for collaboration.

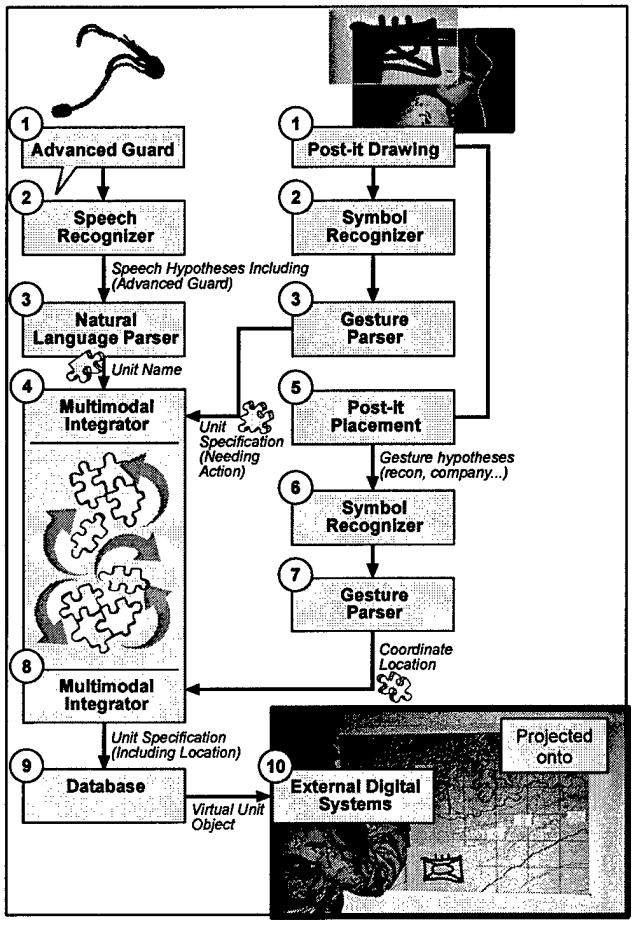

Figure 5. Information flow within Rasa

We have renamed the collection of multimodal components that are outlined below and its user interface for multimodal geo-spatial command-and-control from QuickSet to NISMap and we will use that name in our descriptions below.

Methods, Assumptions, and Procedures

This SBIR project aims to build a paper-based multimodal C2 system that offers the benefits of both paper and digital systems. The system is based on Anoto, AB's concept of "digital paper" – plain paper that has been printed over a special pattern, akin to a watermark. The pattern consists of an array of tiny dots arranged in a quasi-regular grid.

The user can write on this paper using the Anoto pen (see Figure 6), which consists of a regular ink pen, a camera in the pen's tip, a computer, and a Bluetooth wireless transceiver. Figure 7 shows a map printed on Anoto paper. Notice the pattern of tiny dots. When the user writes on the Anoto pattern, the pen acquires the identifier for the paper, and determines where it is located. It then takes images of the grid pattern and can determine where on the paper the pen is drawing. As the user draws, the camera sees only the dot pattern, not the map itself. The ink strokes are captured and stored in the pen's memory, and then sent to a host computer by the wireless network. Currently, all stroke information is sent to the Anoto *Paper Lookup Service* (PLS), which passes this information to the third party application assigned to that paper identifier. The third-party server is told where a given set of digital ink strokes was performed, and thus can

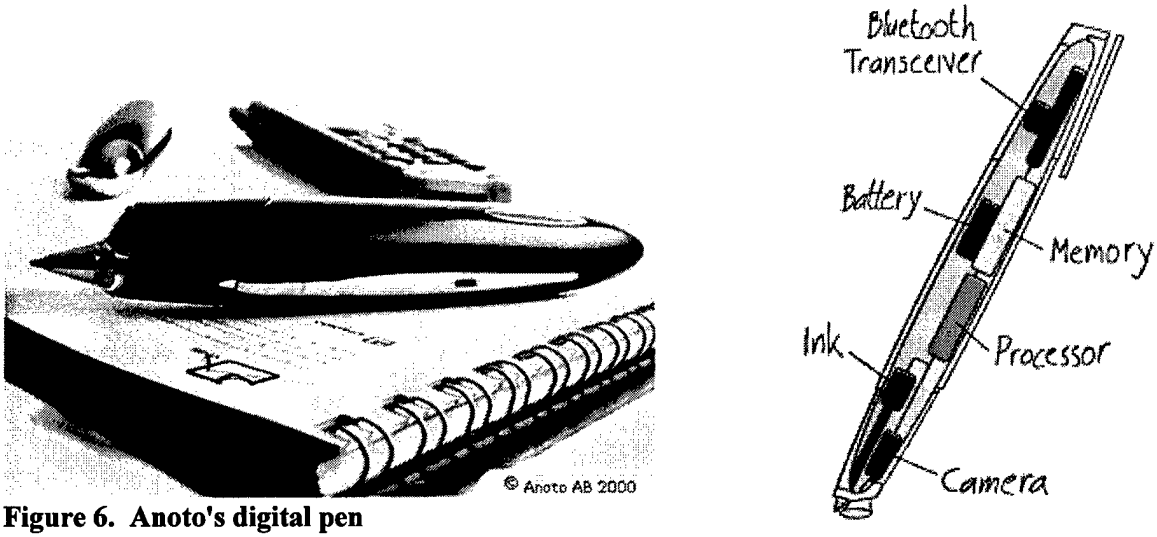

Figure 6. Anoto's digital pen

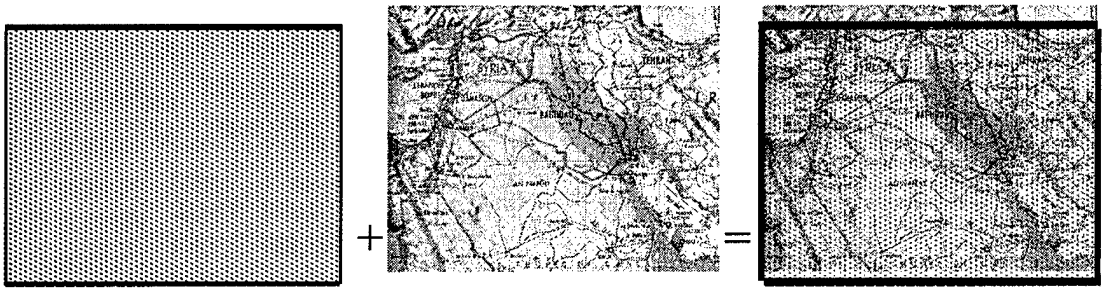

Figure 7: A map printed on Anoto patterned paper

evaluate these strokes of ink to do something intelligent with it. Examples might include recognizing symbology, handwriting, sketches, and standard forms entries such as check boxes, or radio buttons.

In addition to the Anoto grid, which looks like a light gray shading, the paper itself can have standard forms or other patterns, such as a map, printed upon it in non carbon-based inks (see Figure 7). Because the pen saves the data, if the computer or network fails, the

pen can upload the strokes at a later time. Moreover, the paper record is always available.

Because the interface is just paper, it has all the properties one expects of paper – e.g., one can cut/damage the paper, and write on the remainder. Moreover, work can continue if computer systems have failed. This kind of robustness is unusual for computer systems, and will be appreciated by military personnel.

Approach

Our approach for creating a paper-based map prototype system was to develop an agent in our architecture that would respond to the Anoto Paper Lookup Service. Thus, this agent would receive stroke data from the PLS and convert the strokes from a list of generic paper coordinates into a list of geo-spatially oriented coordinates. At this point, the pen input would become identical to that generated by a NISMap client operating on a pen computer, in that the AAA facilitator then routes the strokes to all authorized client programs who have declared an interest in them. In particular, they are sent to the handwriting/symbol recognizer. The text and symbols recognized are then sent back via the facilitator to the NISMap interface, which can project them onto the Anoto paper or place them into the database of other connected systems. Eventually, the raw stroke data, as well as the interpreted symbols/text, are stored in the C2 database. Confirmations, questions, alerts, and reminders can be sent to the officer via wireless LAN (Bluetooth or 802.11b) to be read on his/her PDA, or perhaps on the pen itself (if future versions contain a small screen). So as not to bother the user unnecessarily, the existence of an alert could be indicated through pen vibrations (supported now), or non-speech audio (tones) played in an earphone.

Assuming the user is wearing or is near to a microphone, s/he can also speak to the computer, or to other users. In order to process multimodal inputs the timestamps of the pen need to be coordinated with those of the host computer, which itself needs to be synchronized with the clock of the machine that is processing speech. This overall synchronization is necessary in order that the time-sensitive multimodal integration process can proceed. Once Anoto (or one of their pen licensees) ships a real-time pen, this will be feasible.

Because each pen is uniquely identifiable, and senses its position on the map independently of other pens, multiple users can immediately interact with the same physical map. If each user is wearing a wireless microphone, then each user can reliably generate multimodal interactions simultaneously.

Results and Discussion

During Phase I, we have demonstrated the initial feasibility of a paper-to-computer C^2 interface that will support users' existing work style, yet allow collaboration with users who are employing fully digital systems (see Figure 8). The system developed allows a user to write on a map that has been printed on Anoto paper, and sends the ink strokes to a remote computer system, where they are recognized as symbols and are placed on the

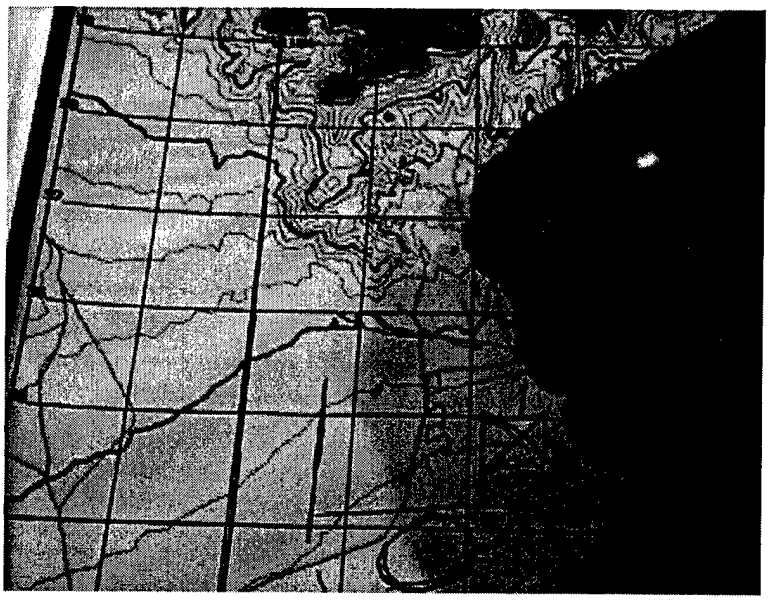

Figure 8. Anoto pen providing input to NISMap

same location on the digital map. Furthermore, it is feasible for NISMap users to work face-to-face on the *same* physical map, a capability unsupported by most digital systems.

Anoto has developed a kit that enables a software developer to overlay his/her own pattern (for us, a map), on the Anoto pattern, have the strokes registered with that overlay, and sent to the developer's application. Based on this kit, NIS completed a sketch-based prototype of its Anoto paper-based command and control application. This prototype enables a user to write symbols on a map that has been printed on Anoto paper. The ink strokes are sent to an agent written for the QuickSet/Rasa architecture that transforms the log file information created by Anoto into the ink representation used by QuickSet and Rasa.[3] The receiving system then recognizes the ink strokes as a symbol, creating the appropriate entity in its database, confirming with both audio and also visually on the QuickSet screen.

During the Phase I extension, we demonstrated the first *multimodal* paper-based C^2 interface based on the Anoto concept by adding speech input to our previous prototype. In order to build it, modifications to the QuickSet/Rasa architecture were made that supported an "open microphone" style of interaction, rather than the Quickset/Rasa

[3] This was necessary because it is implausible that clients would be willing to send their data to Sweden and Anoto has not yet shipped a version of their Enterprise Development Kit, which would allow developers to create their own Paper Lookup Service.

touch-to-talk input style. The touch-to-talk style could not be used because the current Anoto Bluetooth pen from Sony Ericsson operates in batch mode, requiring the user to check a particular box (their Magic Box™) on the paper in order to transmit. However, the QuickSet architecture is time-sensitive—in that gestures and speech must occur within a specified time interval of one another. For a paper-based system, with a batch mode of ink transmission, this timing strategy needed to be changed. Thus, we modified the thresholds to reflect the minimum time that it would take for the pen to establish a Bluetooth transmission to the PC and to convert the ink (about 15-20 seconds). The "open microphone" recognition mode then allows speech and pen to arrive in arbitrary order within the enlarged temporal window. Once the pen includes timing information in the ink strokes, these can be used to disambiguate multiple symbols that may be drawn prior to checking the Magic Box.

Among the other accomplishments was the integration and training of a new symbology recognizer. Many more Mil-Std-2525b symbols can now be recognized, including all echelon boundaries, echelon battle positions, various engineering symbols, ground advance, main attack, screens, and routes. We have also demonstrated recognition of the major classes of unit symbols, both friendly and enemy, including armor, mechanized, infantry, anti-armor, medical, maintenance, reconnaissance, helicopter/airborne, and others. Previously, most of these symbols could only be added multimodally with speech and pointing. Figure 9 depicts a paper sketch demonstrating the recognition of different types of COA entities, such as enemy units, boundaries, and battle positions.

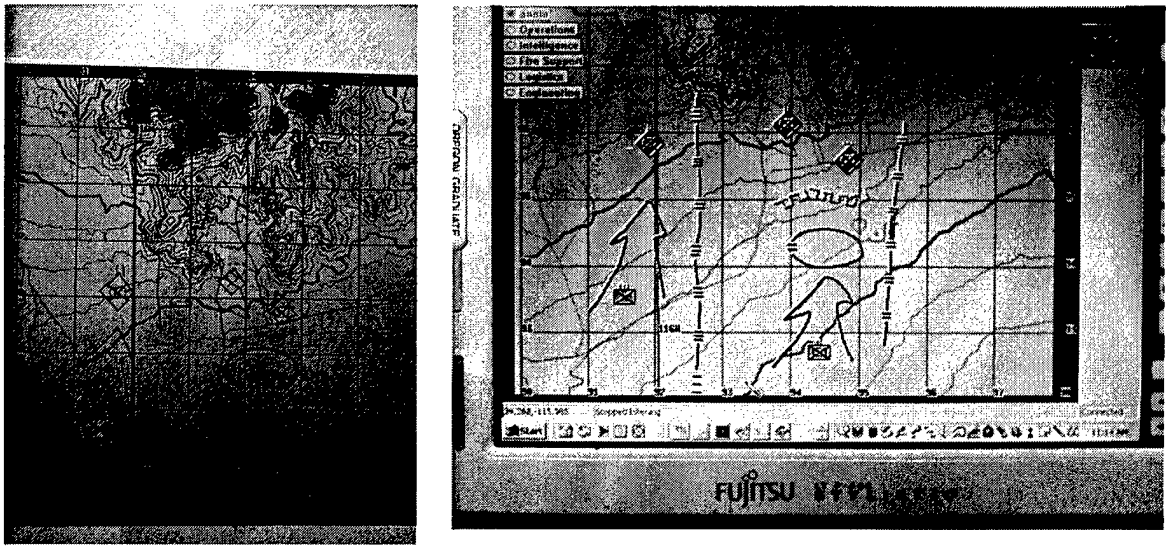

Figure 9: Left --- Paper sketch showing friendly and enemy units, ground advance, boundaries, battle position, and fortification. Right -- System interpretation

Demonstrations

Two demonstrations were performed at each of the following agencies: PM-GCC2, Topographical Engineering Center, and NIMA. As a result, we have obtained a small contract to transition the work to PM-GCC2 via Concurrent Technologies Corporation.

Other Progress

In addition to the primary effort in developing the NISMap prototype, we developed a preliminary prototype application for capturing handwritten notes on shared, collaborative paper displays. Like Rasa, this application would allow people to collaborate via a shared whiteboard-like facility. However, this new application would not be limited strictly to map-based drawings, but would be suitable for more free-form discussions, such as those that colleagues have in their offices, board rooms, and the like.

We made significant progress in designing a proof-of-concept for this application. Notably, the following elements were developed:

- Designed a bitmap for the form and laid out fields on it using our field layout tool.
- Designed and coded an ink exchange file format for ink exchange via shared network drives.
- Designed and coded most of the user interface, including an interface to select one or more ink notes to display simultaneously.
- Reused and adapted components for displaying ink on a bitmap background (useful for projection on a whiteboard as well as simply viewing ink notes).
- Developed a consistent API for collecting ink from either Sony/Ericsson or Logitech pen. Finished research and code to ensure that the same layout would work with Anoto and Logitech pens despite different file format and different coordinate systems.
- Created a DLL out of a handwriting recognizer to use with this application.

Further development was halted due to the release of a 3M Post-it application that had many similarities to the envisioned prototype. The 3M Post-it software application lets users save their Post-it Notes electronically, share them with one another on the network, view a shared bulletin board-like space where digital Post-its can be sorted and changed on their desktop computers. Unlike 3M's application, our prototype would have been able to merge sketches dynamically and in real-time, while collaborators continue to work on paper. There would be support for much larger forms that what is intended for Post-it notes (e.g., SMART Board-sized diagrams). Finally, a variety of backdrops and form layouts would be supported that would encourage the shared work of face-to-face and remote collaborators, working together in concert on such paper diagrams.

Conclusions

During Phase I, we pioneered a new approach that solves both the mobility and the side-by-side collaboration problems. The new system developed during Phase I, called NISMap uses Anoto AB's digital paper and pen, coupled with spoken language, sketch, and other modalities, to enable users to interact multimodally in a 6 ounce form-factor with their favorite paper map products. During the Phase I extension, we incorporated spoken input into this system, thereby developing the first complete prototype of

multimodal interaction with paper. We have also developed a capability to recognize multiple sketched entities from a course-of-action drawing, using spatial properties of ink to segment a batch of strokes into subsets that should be recognized separately. Finally, we have integrated a new, easily trainable symbol/sketch recognizer, and have generated more symbols. An overview of this work and its contribution to computer science will appear in the January issue of the Communications of the ACM and a draft copy of the article accompanies this report.

References

Cohen, P. R., Johnston, M., McGee, D. R., Oviatt, S. L., Pittman, J. A., Smith, I., et al. (1997, November). "QuickSet: multimodal interaction for distributed applications" in the *Proceedings of the International Multimedia Conference*, Seattle, WA, pp. 31-40.

Cohen, P. R. and McGee, D. R. (2004). "Tangible multimodal interfaces for safety-critical applications" in *Communications of the ACM*, Jan. 2004, (special issue on "Conversational Interfaces").

Gorman, P., Ash, J., Lavelle, M., Lyman, J., Delcambre, L., & Maier, D. (2000). Bundles in the Wild: Managing Information to Solve Problems and Maintain Situation Awareness. *Library Trends, 49* (2), pp. 266-289.

Kumar, S., Cohen, P. R., & Levesque, H. J. (2000, July 7-12). "The Adaptive Agent Architecture: Achieving Fault-Tolerance Using Persistent Broker Teams" in the *Proceedings of the International Conference on Multi-Agent Systems*, Boston, MA.

Mackay, W. E. (1999). Is paper safer? The role of flight strips in air traffic control. *ACM Transactions on Computer-Human Interaction, 6* (4), pp. 311-340.

McGee, D. R., Cohen, P. R., & Oviatt, S. L. (1998, August). "Confirmation in multimodal systems" in the *Proceedings of the International Joint Conference of the Association for Computational Linguistics and the International Committee on Computational Linguistics*, Montreal, Quebec, Canada, pp. 823-829.

McGee, D. R., Cohen, P. R., & Wu, L. (2000, April 12-14). "Something from nothing: Augmenting a paper-based work practice with multimodal interaction" in the *Proceedings of the Conference on Designing Augmented Reality Environments*, Helsingor, Denmark, pp. 71-80.

McGee, D. R., Cohen, P. R., Wesson, R. M., & Horman, S. (2002, Apr. 20-25). "Comparing paper and tangible, multimodal tools" in the *Proceedings of the Conference on Human Factors in Computing Systems (CHI'02)*, Minneapolis, MI, pp. 407-414.

Oviatt, S. L. (1996). "Multimodal interfaces for dynamic interactive maps" in the *Proceedings of the Conference on Human Factors in Computing Systems*, pp. 95-102.

Oviatt, S. L. (1999). "Mutual disambiguation of recognition errors in a multimodal architecture" in the *Proceedings of the Conference on Human Factors in Computing Systems*, Pittsburgh, PA, pp. 576-583.

Oviatt, S. L. (2000a). "Taming speech recognition errors within a multimodal interface" in *Communications of the ACM*, Sept. 2000, 43 (9), 45-51 (special issue on "Conversational Interfaces").

Oviatt, S. L., (2000b) "Multimodal system processing in mobile environments," in the *Proceedings of the Thirteenth Annual ACM Symposium on User Interface Software Technology (UIST'2000)*, 21-30. New York: ACM Press.

Distribution List

Commander — 2 Copy
U.S. Army Aviation and Missile Command
ATTN: AMSAM-RD-WS-DP-SB (Mr. Alexander H. Roach, Tech Monitor)
Bldg 7804, Room 205
Redstone Arsenal, AL 35898

Commander — 1 Copy
U.S. Army Aviation and Missile Command
ATTN: AMSAM-RD-OB-R
Bldg 4484, Room 204
Redstone Arsenal, AL 35898-5241

Commander — 1 Copy
U.S. Army Aviation and Missile Command
ATTN: AMSAM-RD-WS
Bldg 7804, Room 247
Redstone Arsenal, AL 35898-5248

Director — 1 Copy
Defense Advanced Research Projects Agency
ATTN: IXO (Mr. Ward Page)
701 North Fairfax Drive
Arlington, VA 22203-1714

Director — 1 Copy
Defense Advanced Research Projects Agency
ATTN: IPTO (LCDR Dylan Schmorrow)
701 North Fairfax Drive
Arlington, VA 22203-1714

Director — 1 Copy
Defense Advanced Research Projects Agency
ATTN CMO/SBIR
3701 North Fairfax Drive
Arlington, VA 22203-1714

Director — 1 Copy
Defense Advanced Research Projects Agency
ATTN: OMO/DARPA Library
3701 North Fairfax Drive
Arlington, VA 22203-1714

Defense Technical Information Center — 2 Copies
ATTN: Acquisitions/DTIC-OCP, Rm-815
8725 John J. Kingman Rd., STE 0944
Ft. Belvoir, VA 22060-6218

DRAFT: Not for distribution or quotation. To appear in *Communications of the ACM*, January, 2004.

Tangible Multimodal Interfaces for Safety-Critical Applications

Philip R. Cohen and David R. McGee*
Natural Interaction Systems, LLC
{pcohen,dmcgee}@naturalinteraction.com

Introduction

Despite the success of information technology, there are important problem-solving tasks that computing has simply failed to assist. Consider an example from the military. In Figure 1, we see officers turning their backs on computing, preferring instead to work with an 8-foot high paper map and Post-it™ notes. They explain that there are good reasons for their reluctance to use digital systems — paper maps are readily available, lightweight, cheap, high in resolution, very large or small in scale, supportive of collaboration, and fail-safe. Likewise, a trip to most local hospitals should convince the observer that physicians have yet to abandon paper despite the advantages of electronic medical records (Gorman et al., 2000). Air traffic controllers have a similar preference for paper "flight strips" in performing their stressful jobs (Mackay et al., 1999).

Over time professionals in these safety-critical domains have developed manual work practices that their organizations are attempting to replace with computer systems. Though there are good reasons to automate, and the resulting systems may have been built with the prevailing best practices, they nonetheless alter fundamental aspects of what users do and value. Not surprisingly, the systems encounter resistance. For example, at Cedars Sinai hospital in Los Angeles, physicians rebelled against the installation of a physician order entry system, causing it to be removed:

> "I'm not opposed to change ... but it's got to be new and better," said Dudley Danoff, MD, a urologic surgeon who helped organize physician opposition. "This was new but certainly not better" than paper... Cedars-Sinai's decision was extraordinary but not unique. David Classen, MD, of First Consulting Group, says he knows of at least six other hospitals that have pulled paperless systems in the face of physician resistance and other problems. (American Medical News, Feb. 17, 2003. http://www.ama-assn.org/scipubs/amnews/pick_03/bil20217.htm).

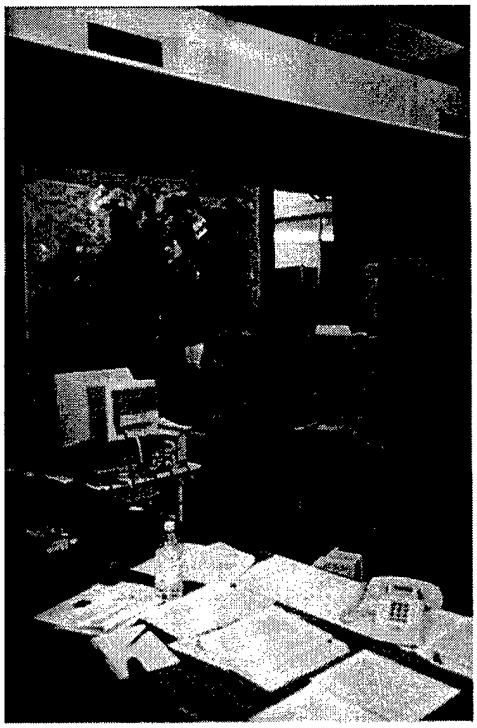

Figure 1. Officers tracking a battle. Photo courtesy of William Scherlis.

These failures are predicted by Moore's analysis of technology acceptance (Moore 1991). According to Moore, a chasm exists between early adopters and a much larger group, comprised of so-called "pragmatists" and "conservatives," in how each group tends to accept technology products. Whereas the former are willing to expend energy on learning and promoting revolutionary technology, the latter prefer evolution over revolution. They want technology that enhances and integrates easily with existing work practices and systems, and that has been used successfully by their peers.

One reason digital systems have fallen into this chasm is that computing hardware and its human interfaces do not fit the way these professionals work. In particular, laptop, keyboard, mice, trackball, and similar interaction devices are not optimal for field and mobile use, or for face-to-face collaboration. Pen-based tablets and PDAs address some users' needs for mobility and ease of use, but these systems need far better resolution, much less weight, and better portability. Moreover, they are currently limited by their reliance on the desktop metaphor for graphical user interfaces (GUIs), which focuses users' attention on the computer itself and on overlapping windows, menus, files and folders. This distraction from the task can be counterproductive when providing health care, planning battles, or guiding aircraft. Finally, because work stoppages can be costly or life-threatening, users in these safety-critical domains remain concerned about system or network crashes.

Rather than require that users change, system designers could adapt their systems to key aspects of the users' work practice. Indeed, *tangible multimodal systems (TMMs)*, described below, enable users to employ physical objects already in their workplace (e.g., paper or other physical tools), along with natural spoken language, sketch, gesture, and other input modalities to interact with information and with coworkers. In this article, we discuss tangible and multimodal interfaces separately, and then illustrate how they can be combined to produce TMMs.

Bridging the Chasm with Tangible Multimodal Interfaces

Tangible user interfaces

Tangible user interfaces (TUIs) incorporate physical objects as sensors and effectors that, when manipulated, modify computational behavior. To enable the construction of TUIs, systems distinguish and identify physical objects, determine their location, orientation, or other aspects of their physical state, support annotations on them, and associate them with different computational states. To do this, TUIs use technologies such as radio emitters, bar codes, or computer vision. Wellner developed the first tangible paper system, the DigitalDesk (Wellner, 1993), incorporating paper via computer vision. The DigitalDesk could copy and paste printed text or numbers from paper into digital documents via OCR, enabling the user to manipulate the information electronically. Mackay and her collaborators have explored tangible flight-strip prototypes for the air traffic control industry since the early '90s (Mackay, et al. 1999). Their prototypes captured handwritten annotations on paper strips, and tracked the strips' relative locations on a mounting board using video capture techniques and electrical resistance. Ishii and his students in the MIT Media Laboratory also have developed various tangible prototypes, notably the Urp system (Underkoffler & Ishii, 1999), which again used video tracking to support the use of physical tools, such as rulers and clocks, within an urban planning setting.

Multimodal interfaces

With flexible multimodal interfaces users can take advantage of more than one of their natural communication modes during human-computer interaction, selecting the best mode or combination of modes that suit their situation and task. For example, the QuickSet multimodal system enables a user to speak while sketching (Cohen et al., 1997; Oviatt and Cohen, 2000). Here, sketch provides spatially-oriented information such as shape and location, while speech provides information about identity, speed, color, and other attributes. Multimodal interfaces scale to very large or small devices, and can be used within sensor-rich environments. Another important benefit that derives from the ability of an interface to support use of multiple modalities is *mutual disambiguation*, in which information provided by one or more sources can be used to resolve ambiguities in another, thereby reducing errors (Oviatt, 1999). Thus, multimodal systems are better equipped to manage the inherent uncertainty of sensors and recognizers than are systems that rely only on a single uncertain information source. Furthermore they enable more efficient performance of various tasks—research has found multimodal interfaces to be four- to nine-fold faster for map-

based tasks compared to GUIs, with no increase in the number of errors (Cohen et al., 2000).

Combining TUIs and MMUIs

Although tangible systems allow users to manipulate physical objects, they typically do not interpret any annotations that accompany those objects. On the other hand, multimodal interfaces can analyze users' verbal and written information, but they typically do not acquire information from objects situated in the real world. Thus, neither individual interface style is as well equipped to satisfy the workplace needs of the applications discussed earlier as is the combination of those styles. As a guide to designing such tangible multimodal systems for safety-critical situations, we have argued (McGee, et al. 2000) that systems need to meet the following constraints:

- *Minimality Constraint:* Changes to users' work practice should be minimized; computational tools should be based on the physical tools, procedures, and language already employed by the users.
- *Human Performance Constraint:* Users should be able to annotate and manipulate the physical objects in their work environment easily, in order to accomplish tasks.
- *Human* Understanding *Constraint*: At any time, users should be able to understand what each physical artifact represents, both computationally and in the real world.
- *Malleability Constraint*: Because situations can change, users must be able to modify the meanings they attribute to physical artifacts.
- *Robustness Constraint:* Should the system, communications, or power fail, users must still be able to understand the situation and continue to interact with the physical objects without interruption.

In order to build systems that satisfy these constraints, it is important to notice that the work practice routines employed by these professionals typically impart meaning to the physical objects in their environments using a shared workplace-dependent language and/or symbology. If a system could support both physically and computationally these existing workplace routines and languages, it could meet the minimality, human performance, malleability, and human understanding constraints discussed above. Such a system would permit the users to be able to create, change, and understand the meanings associated with the workplace's physical objects in a familiar fashion, while simultaneously updating a digital version of the information. Tangible multimodal systems meet all of these requirements by[1]:

- Processing the relevant state changes of physical work objects;
- Understanding users' task-related multimodal communication,
- Fusing information from physical, linguistic, gestural, and other information sources, thereby managing uncertainty and recognition errors;
- Delivering relevant information and confirmation to users in a manner that is appropriately integrated with the physical work environment.

We have developed three systems, *Rasa,*™ *NISMap,*™ and *NISChart,*™ that demonstrate the feasibility of TMMs.

Rasa enables a military officer to use paper maps and Post-It™ notes (McGee, et al. 2000) in support of command and control tasks. During battle tracking, officers plot unit locations, activities, and other elements on a paper map by drawing unit symbols on Post-It notes, positioning them on the map, and then moving them in response to incoming reports of their new locations. With Rasa (Figure 2), each of the pieces of paper is mounted on a digitizing tablet — the map is registered to a large touch-sensitive tablet, and the Post-Its rest upon a tablet that supports both digital and physical ink. The user writes a military unit symbol (e.g., for an armored platoon) on a Post-It note. The user can also speak identifying information about that unit (e.g., "First platoon, Charlie Company") while drawing the symbol. The computer recognizes both the symbol and the voice, fusing their meanings using multimodal integration techniques (Cohen, et al.

[1] See (McGee et al., 2000) for a detailed example.

1997). Then, the user places the Post-It onto the paper map, which causes the unit that was drawn (and sometimes spoken about) to be recorded in the system's database, with its location specified by the place it was mounted on the map. The system projects a unit symbol onto the relevant location on the paper map, and distributes the result to collaborating systems. If a report arrives indicating that the unit has moved, the user only needs to pick up the note and put it at the new location on the map.

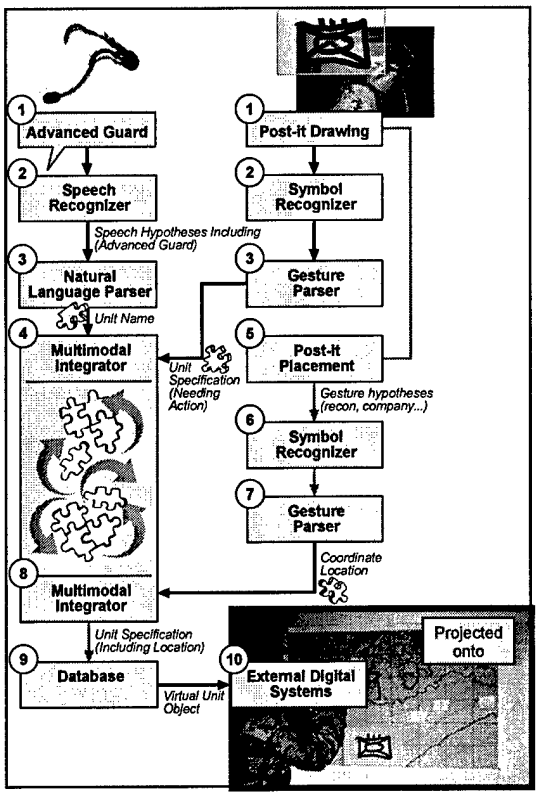

Figure 2. Information flow within Rasa

The robustness constraint dictates that work must somehow continue during a system or communications failure. What happens if the computer supporting Rasa goes down? In order to investigate this question, we undertook an experiment (McGee et al., 2002) in which officers were studied as they tracked a battle. During each session, Rasa was deliberately "crashed," but reports kept arriving. In response, officers simply continued to create and move the Post-It notes on the paper map. When the computer came back online, it digitally projected the old unit locations onto the paper map, whereas the notes indicated the units' current locations. It was then a simple process to reconcile the computer system with the paper version. Thus, because the physical objects constituted the user interface, no additional ongoing "backup" process was needed.

Whereas Rasa enables users to employ paper maps in performing their tasks, the maps still require large, relatively immobile digitizers. Moreover, because of limitations in current digitizing tablets, only one person can write on the map at a time. Thus, Rasa is still limited in its mobility and support for collaboration. We realized early on that Rasa was a prototype of a much larger set of techniques for TMMs, wherein paper is the principle component of the work practice. We call these techniques *Multimodal Interaction with Paper* (*MIP*).

Our new MIP applications employ Sweden-based Anoto AB's digital pen and paper, along with Rasa's multimodal processing capabilities. In these applications, the paper has a dot-pattern printed on it, and over the dots is printed content information (e.g., a map). The Anoto pen produces ink like any other pen, but it also has a Strong ARM CPU, memory, a Bluetooth radio, and a camera that sees only the dot pattern (Figure 3). The pen decodes the locations it has observed, saving them in memory and/or transmitting them wirelessly via Bluetooth or via a USB-linked "ink well." With the Bluetooth version of MIP, a user can be 30ft. away from the receiving computer and can draw on an ordinary piece of paper that has been covered with the Anoto dot pattern.

NISMap. Like Rasa, the NISMap user can speak and/or sketch on a paper map (Figure 3). In response, the system collects the user's strokes, recognizes writing and/or symbols, correlates and fuses co-temporal speech, and updates a central database serving other systems and colleagues. Because this TMM has the portability, high resolution, scalability, and physical properties of pen and paper, it meets the needs of officers in the field, in particular, robustness to computer failure. Furthermore, multiple users can write on the same map at the same time, so this system provides unique support for face-to-face collaboration. Finally,

NISMap addresses officers' concerns that a computer map with a hole in it is a "rock," while a paper map with a hole in it is still a paper map—NISMap continues to work even if the paper has been crumpled, punctured, torn, or taped up.

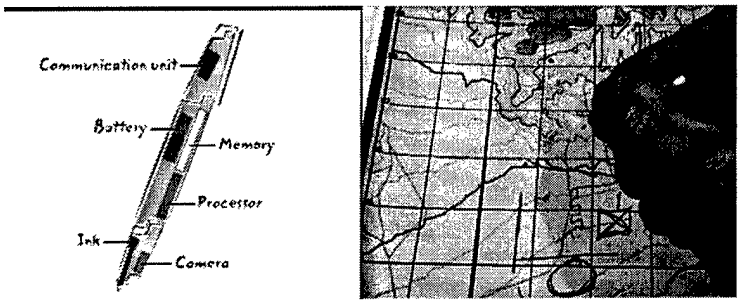

Figure 3. Left: Digital pen enabling Anoto functionality; Right: NISMap application using Anoto pen and paper

NISChart. Physicians are accustomed to paper forms. However, if there is only a paper record, opportunities are missed for improving both individual care and institutional efficiency. Whereas the computer-based patient record is identified by the Institute of Medicine as a requirement for better medical care, they caution: "Perhaps the single greatest challenge that has consistently confronted every clinical system developer is to engage clinicians in direct data entry" (IOM, p. 125). They claim that "To make it simple for the practitioner to interact with the record, data entry must be almost as easy as writing." (ibid. p. 88) In accord with this suggestion, we have built NISChart, a digital-paper based charting solution where writing on paper and speaking are the primary input modalities. The system allows a physician to enter values, text, check marks, etc. into the hospital's standard forms, printed on Anoto paper. Digital ink is transmitted to the application, which applies contextual and semantic knowledge in conjunction with handwriting, symbol, and speech recognition to populate a relational database. The information is stored in its digital form, either as traditional database entries (e.g., text and symbols) or as digital ink. NISChart provides graphical and/or verbal feedback at the point of data entry or to other workers. Finally, the physical paper with real ink can serve as the definitive primary record, which is important both for recovering from failure and for legal reasons. NISChart is currently being alpha tested at a major urban medical center.

Conclusion

We have argued that computing too often requires professionals to alter their work to fit current technology. One critical area in which this mismatch occurs is the user interface. Because most professionals value their skills, time, and ability to interact with clients or colleagues, they resist the introduction of computer systems that do not treat those values as paramount. In order to reach these skilled professionals, we suggest that safety-critical systems let users continue to employ the physical objects, language, and symbology of their workplace through the use of tangible multimodal systems. TMM users are as capable of updating digital systems and of collaborating digitally with colleagues as are the users of more traditional systems. At the same time, TMM-based digital systems benefit from many features of the physical world, such as those of paper. Ultimately, rather than being stigmatized as "late adopters" of systems, users of tangible multimodal interfaces could be included among the "power users" of next-generation technologies, while at the same time reaping benefits that users of more traditional digital systems lack.

References

Cohen, P. R., Johnston, M., McGee, D. R., Oviatt, S. L., Pittman, J. A., Smith, I., et al. (1997, November). "QuickSet: multimodal interaction for distributed applications" in the *Proceedings of the IEEE International Multimedia Conference*, Seattle, WA, pp. 31-40.

Cohen, P. R., McGee, D. R., & Clow, J. (2000, April). "The efficiency of multimodal interaction for a map-based task" in

the *Proceedings of the Applied Natural Language Programming Conference*, Seattle, WA, pp. 331-338.

Gorman, P., Ash, J., Lavelle, M., Lyman, J., Delcambre, L., & Maier, D. (2000). "Bundles in the wild: Managing information to solve problems and maintain situation awareness." *Library Trends, 49* (2), pp. 266-289.

Institute of Medicine. (1997). *The Computer-based Patient Record: An Essential Technology for Health Care, 2nd edition*, National Academy Press.

Mackay, W. E. (1999). "Is paper safer? The role of flight strips in air traffic control." *ACM Transactions on Computer-Human Interaction, 6* (4), pp. 311-340.

McGee, D. R., Cohen, P. R., Wesson, R. M., & Horman, S. (2002, Apr. 20-25). "Comparing paper and tangible multimodal tools" in the *Proceedings of the Conference on Human Factors in Computing Systems (CHI'02)*, Minneapolis, MI, pp. 407-414.

McGee, D. R., Cohen, P. R., & Wu, L. (2000, April 12-14). "Something from nothing: Augmenting a paper-based work practice with multimodal interaction" in the *Proceedings of the Conference on Designing Augmented Reality Environments*, Helsingor, Denmark, pp. 71-80.

Moore, G. A. (1991). *Crossing the Chasm: Marketing and Selling High-Tech Goods to Mainstream Customers*. New York: Harper Business.

Oviatt, S. L. "Mutual disambiguation of recognition errors in a multimodal architecture," *Proceedings of Conference on Human Factors in Computing Systems: CHI '99,* New York, N.Y.: ACM Press, 1999, pp. 576-583.

Oviatt, S.L. & Cohen, P.R. (2000) "Multimodal Interfaces That Process What Comes Naturally" *Communications of the ACM*, Vol. 43, No. 3, March, pp. 45-53.

Underkoffler, J. & Ishii, H. "Urp: A luminous tangible workbench for urban planning and design," in the Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI'99), May 1999, ACM Press, pp. 386-393.

Wellner, P. (1993). "Interacting with paper on the DigitalDesk." *Communications of the ACM, 36* (7), pp. 87-96.

---

* The research discussed here has been funded partly by DARPA contract N66001-99-D-8503 to the Oregon Graduate Institute of Science and Technology, part of the Oregon Health and Science University, where Cohen is a faculty member and McGee received his Ph.D. The research was also partly supported by DARPA SBIR contract DAAH0102CR051 to Natural Interaction Systems, LLC. However, the views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Defense Advanced Research Projects Agency or the U.S. Government.

